AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

ENVIRONMENTAL MICROBIOLOGY

# Long-Range PCR Reveals the Genetic Cargo of IncP-1 Plasmids in the Complex Microbial Community of an On-Farm Biopurification System Treating Pesticide-Contaminated Wastewater

**Vincent Dunon,[a] Peter N. Holmsgaard,[b] Simone Dealtry,[c] Rob Lavigne,[d] Søren J. Sørensen,[b] Kornelia Smalla,[c] Eva M. Top,[e] Dirk Springael[a]**

[a]Division of Soil and Water Management, KU Leuven, Heverlee, Belgium
[b]Department of Biology, Section of Microbiology, University of Copenhagen, Copenhagen, Denmark
[c]Julius Kühn-Institut (JKI) – Federal Research Centre for Cultivated Plants, Dept. Epidemiology and Pathogen Diagnostics, Braunschweig, Germany
[d]Division of Gene Technology, KU Leuven, Heverlee, Belgium
[e]Department of Biological Sciences, Institute for Bioinformatics and Evolutionary Studies (IBEST), University of Idaho, Moscow, Idaho, USA

**ABSTRACT** Promiscuous plasmids like IncP-1 plasmids play an important role in the bacterial adaptation to pollution by acquiring and distributing xenobiotic catabolic genes. However, most information comes from isolates and the role of plasmids in governing community-wide bacterial adaptation to xenobiotics and other adaptive forces is not fully understood. Current information on the contribution of IncP-1 plasmids in community adaptation is limited because methods are lacking that directly isolate and identify the plasmid borne adaptive functions in whole-community DNA. In this study, we optimized long-range PCR to directly access and identify the cargo carried by IncP-1 plasmids in environmental DNA. The DNA between the IncP-1 backbone genes *trbP* and *traC*, a main insertion site of adaptive trait determinants, is amplified and its content analyzed by high-throughput sequencing. The method was applied to DNA of an on-farm biopurification system (BPS), treating pesticide contaminated wastewater, to examine whether horizontal gene exchange of catabolic functions by IncP-1 plasmids is a main driver of community adaptation in BPS. The cargo recovered from BPS community DNA encoded catabolic but also resistance traits and various other (un)known functions. Unexpectedly, genes with catabolic traits composed only a minor fraction of the cargo, indicating that the IncP-1 region between *trbP* and *traC* is not a major contributor to catabolic adaptation of the BPS microbiome. Instead, it contains a functionally diverse set of genes which either may assist biodegradation functions, be remnants of random gene recruitment, or confer other crucial functions for proliferation in the BPS environment.

**IMPORTANCE** This study presents a long-range PCR for direct and cultivation-independent access to the identity of the cargo of a major insertion hot spot of adaptive genes in IncP-1 plasmids and hence a new mobilome tool for understanding the role of IncP-1 plasmids in complex communities. The method was applied to DNA of an on-farm biopurification system (BPS) treating pesticide-contaminated wastewater, aiming at new insights on whether horizontal exchange of catabolic functions by IncP-1 plasmids is a main driver of community adaptation in BPS. Unexpectedly, catabolic functions represented a small fraction of the cargo genes while multiple other gene functions were recovered. These results show that the cargo of the target insertion hot spot in IncP-1 plasmids in a community, not necessarily relates to the main obvious selective trait imposed on that community. Instead, these functions might contribute to adaptation to unknown selective forces or represent remnants of random gene recruitment.

**Editor** Maia Kivisaar, University of Tartu

Address correspondence to Dirk Springael, dirk.springael@kuleuven.be.
The authors declare no conflict of interest.

**Received** 17 August 2021
**Accepted** 25 November 2021
**Accepted manuscript posted online** 8 December 2021
**Published** 8 February 2022

**KEYWORDS** horizontal gene transfer, IncP-1 plasmid, adaptive traits, biopurification system, catabolic genes, genetic cargo, microbial community

Horizontal gene transfer (HGT) plays an important role in bacterial adaptation to organic xenobiotics by promoting the dissemination of catabolic genes between different community members. This can accelerate the distribution and even *in vivo* construction of novel metabolic pathways in bacteria (1–3). HGT is mediated by mobile genetic elements (MGEs) and the collective pool of MGEs in an environment is referred to as the mobilome. The mobilome includes MGEs such as (conjugative) plasmids, transposons, integrons and insertion sequences (IS). A group of mobile elements that appear key in catabolic adaptation of environmental communities is the group of the IncP-1 plasmids. IncP-1 plasmids are known for their broad host range, their highly promiscuous nature and their often large accessory gene regions and hence high potential to promote the intra- and interspecies exchange of genetic material (4, 5). The accessory regions carry the genes that ensure the adaptation and survival of bacteria to their local environment such as the exposure to abiotic and biotic stressors. For instance, in organic xenobiotic-degrading bacteria, the catabolic genes are often cargo of IncP-1 plasmids (6, 7). The importance of IncP-1 plasmids in community adaptation in contaminated environmental settings is further suggested by their high abundances and diversity in several of such settings as revealed by canonical or quantitative PCR (8, 9). However, the plasmid-related processes that govern adaptation in complex communities are not yet fully understood. Especially, key knowledge gaps exist about the nature of the cargo genes that such plasmids carry and distribute in complex communities. Indeed, as for other plasmids, current information on the contribution of IncP-1 plasmids in community adaptation in a certain environment is mainly based on isolates originating from that environment or occasionally from exogenous plasmid isolation (10–12). These approaches are prone to biases since they involve cultivation and selection for specific traits, thus introducing a bias in the cargo of the plasmids toward that trait which in reality might only represent a minor fraction of the diversity (10). To acquire a holistic community-wide picture of the identity of the IncP-1 cargo genes and how gene acquisition by IncP-1 plasmids affects community diversification, cultivation-independent analysis techniques are required that directly indicate the plasmid borne adaptive functions in community DNA. However, such methods are currently lacking. Approaches like shot-gun metagenomics are often unsuitable to link gene function with plasmid vectors, especially when it concerns rare populations within the community (13). Alternatively, community-wide plasmid isolation and sequencing can be applied but recovering larger plasmids like IncP-1 plasmids, comes often with difficulties (14). Clearly, the availability of cultivation-independent methods that identify cargo genes of IncP-1 plasmids in community DNA would add greatly to our understanding of the role of IncP-1 plasmids in microbial community adaptation in contaminated environments as well as in other ecosystems.

Long-range PCR (LR-PCR) refers to the amplification of long DNA fragments that cannot typically be amplified by routine PCR methods or reagents. It is mainly used for cloning larger DNA fragments, genome mapping and sequencing and contig construction (15). However, LR-PCR also provides opportunities for targeted culture-independent metagenomic analysis of microbial communities in ecology-oriented studies (16–18). Indeed, when combining LR-PCR with high-throughput sequencing of the amplicons, it allows us to enrich for and collect sequence information on specific subsets of the genetic material in an environment. This is of particular interest for highly diverse microbial ecosystems with potentially small fractions of functionally relevant bacteria, which could be easily missed by traditional shotgun sequencing. In this study, in analogy to the method that we previously reported to capture and identify IS*1071* associated cargo from environmental DNA (eDNA) (16), we here present a LR-PCR method to recover and sequence cargo genes carried by IncP-1 plasmids. The method makes use of the highly conserved modular structure of IncP-1 plasmids (Supplemental material

Fig. S1) characterized by two distinct hot spots for insertion of cargo genes in the IncP-1 backbone, i.e., between the *trb* and *tra*-operon (accessory region 1) and between *trfA* and the origin of replication (*oriV*) (accessory region 2) (17). These predictable locations of cargo genes on IncP-1 plasmids allows designing primers that anneal to the conserved backbone gene sequences flanking the cargo in LR-PCR which then will specifically amplify the cargo region. In this study, we focused on accessory region 1 located between the *trb* and *tra*-operon. The novel LR-PCR assay was used to examine the question which gene functions are actually carried within accessory gene region 1 of IncP-1 plasmids hosted by the microbial community of an on-farm biopurification system (BPS). BPS treat pesticide contaminated wastewater originating from filling and cleansing of pesticide spraying equipment at the farm yard or from postharvest processing of agricultural produce (18). They are considered a time-, labor- and cost-efficient technology to control pesticide point source contamination of surface waters. In the BPS, the wastewater is filtered through a so-called biomix consisting of soil and organic waste. The pesticides are either removed by sorption and/or biodegradation, of which biodegradation to innocuous products is the most preferred (19). Microbial degradation has indeed been observed in BPSs, but the mechanisms involved in its establishment are poorly understood (18). Given the high pesticide loading rates of BPSs, it can be hypothesized that the BPS environment provides highly selective conditions for *in vivo* genetic adaptation of microbiota toward catabolism of pesticides and proliferation of pesticide degrading bacteria involving HGT (8, 16). Several recent studies suggest that IncP-1 plasmids are implicated in community adaptation in the BPS environment. Abundances of IncP-1 marker genes are exceptionally high in the BPS matrix (8, 20, 21). Moreover, recently we showed the high incidence of xenobiotic catabolic genes as cargo of composite transposons flanked by the IS element IS*1071* in a BPS environment (16). Since IS*1071* associated composite transposons carrying xenobiotic catabolic genes are often located on IncP-1 plasmids in isolates, their high incidence in the BPS environment might be due to recruitment by IncP-1 plasmids. The examined BPS was a long-term operational (>3 years) BPS located on a farm in Kortrijk, Belgium (BPS Kortrijk), with analyzed samples taken over the course of the agricultural season of 2011. The same BPS and samples were previously scrutinized regarding their community composition, IncP-1 diversity, and cargo of IS*1071* flanked composite transposons (16, 21–23). Given the high and continuous pesticide exposure of a BPS, we hypothesized that the recovered LR-PCR amplicons, and hence accessory region 1 of IncP-1 plasmids hosted by the BPS microbial community, would be strongly enriched in pesticide catabolic genes.

## RESULTS

**Optimization of LR-PCR to recover IncP-1 plasmid accessory region 1 genes from eDNA.** To amplify accessory region 1 of IncP-1 plasmids, the backbone genes *traC* and *trbP* were chosen as targets for the LR-PCR primers. As detailed in Supplementary information, these two genes were the most ubiquitous flanking backbone genes of this accessory region within the used reference IncP-1 plasmid set. In several plasmids, *trbP* not always directly flanked accessory region 1 as other backbone genes like *parA*, *upf30.5* and *upf31.0* intervened between *trbP* and the accessory region meaning that these genes might be co-amplified. As the sequences of *trbP* varied between different IncP-1 subgroups, a consensus primer set that targets the accessory region 1 could only be designed for the subgroups $\beta$, $\varepsilon$ and $\delta$, and not for the subgroups $\alpha$, $\gamma$, and $\zeta$ which are though much less frequently found in environmental habitats. Details about the matching of the selected primer sequences with target and nontarget plasmids are further reported in Supplementary information. The LR-PCR was tested on IncP-1 reference plasmids belonging to the main five IncP-1 subgroups -$\alpha$, -$\beta$, -$\gamma$, -$\delta$, and -$\varepsilon$ (Fig. S2). As intended, amplification of the expected fragment was successful for the target subgroups $\beta$, $\varepsilon$ and $\delta$ (including pAKD4) (see green arrows Fig. S2 and Fig. 1). The cutoff value of amplicon size appears to be around 21 kb since no amplicons were obtained when using template DNA of plasmids containing accessory 1 regions longer than 20 kb,

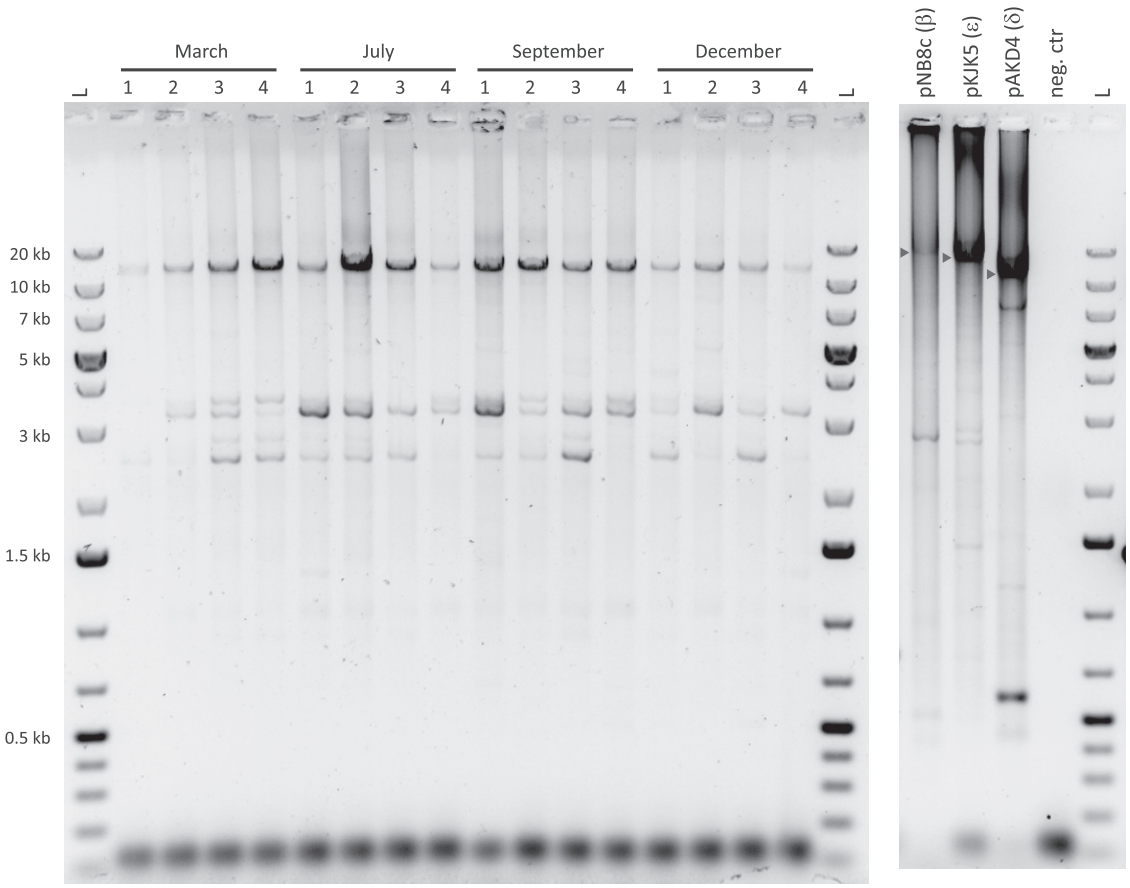


**FIG 1** Agarose gel electrophoresis of IncP-1 accessory region 1 LR-PCR amplicons recovered from BPS Kortrijk samples taken in the course of the agricultural season of 2011. Samples from BPS Kortrijk were taken in March, July, September and December. At each time point. high molecular weight eDNA was isolated from four samples taken at four different positions (compartments 1–4) of the BPS and used as the template for LR-PCR. Plasmid DNA of IncP-1 plasmids pNB8c ($\beta$), pKJK5 ($\varepsilon$) and pAKD4 ($\delta$) were included as positive controls with expected amplicon sizes of, respectively, 20.9, 17.2 and 12.3 kb (red arrows). A nontemplate control was included as negative control (neg. ctr). To estimate amplicon sizes a lane (L) containing GeneRuler$^{TM}$ 1 kb plus DNA ladder was included.

even when increasing the elongation time. Small unexpected amplification products, though, were observed with several plasmids (see red and yellow arrows Fig. S2). As outlined in the Supplementary information, many of these PCR products, i.e., the amplicons indicated by red arrows in Fig. S2, could be explained by the presence of an 8- to 12-bp sequence within *traC* that matches the 3′ end of the trbP_LR primer sequence gene, upstream of the traC_LR primer annealing sequence. Other small amplicons (indicated in green in Fig. S2), except for an around 800 bp amplicon recovered with pI2 and pTB30, could be explained by 3′ end trbP_LR primer matching sequences in other genes located just upstream of *traC, including* a 8 bp matching sequence in IS*1071* (Table S2). No suitable primers could be found that avoided the amplification of these extra internal plasmid fragments. Since (i) the amplification of these fragments did not hamper the amplification of accessory region 1, (ii) the traC_LR primer proofed to be highly specific and hence only a fragment can be amplified that extends from *traC* in the direction of *trbP*, and (iii) the majority of these amplicons can be easily explained and recognized by agarose gel electrophoresis and sequencing, it was decided to continue with the trbP_LR and traC_LR primer set. As a proof of concept to recover IncP-1 accessory genes from eDNA and to determine the detection limit, the trbP_LR/traC_LR primer set was used in LR-PCR on eDNA isolated from BPS Kortrijk material that was spiked with a 10-fold dilution series of *E. coli* containing plasmid pNB8c. This plasmid carries the catabolic gene cluster *dcaRQTA1A2BC* in accessory region 1. An amplicon was recovered of the expected size of around 21 kb. To confirm the identity of the amplicon, the LR-PCR product was digested

**TABLE 1** Summary of the analysis of the sequencing data obtained from the IncP-1 accessory region 1 LR-PCR amplicons recovered from BPS Kortrijk eDNA[a]

| Description | Value |
|---|---|
| Clean sequencing data (Mbp) | 521 |
| Assembled sequence data (kb) | 499 |
| Contigs | 333 |
| Size range of contigs (kb) | 0.5 – 9.6 |
| N50 (bp) | 1930 |
| Contigs with *traC* end | 5 |
| Contigs with *trb*[b] end | 10 |
| Contigs with *traC* and *trb*[b] end | 1 |
| ORFs | 430 |
| ORFs with KO | 137 |
| ORFs with known function BLASTp (nr) | 274 |
| ORFs with known function BLASTp (UniprotKB) | 233 |
| ORFs with HMM | 279 |

[a]Contigs with IncP-1 remnants and the number of open reading frames (ORFs) were predicted by RAST. Annotations of the ORFs were made by assigning either a KO number using blastKOALA or a known protein function using BLASTp or the retrieval of Hidden Markov Models (HMM) using hmmscan.

[b]Contigs containing (parts of) *trb*P, *upf*30.5, *upf*31.0 or *par*A genes are considered IncP-1 backbone markers for the *trb*-operon insertion site.

with the restriction enzyme AhdI followed by Southern blotting hybridization using *dcaQ* as a DNA probe. The expected hybridization pattern was obtained (data not shown). The detection limit was around $10^5$ copies of pNB8c per gram of sample material (data not shown).

**Diversity of accessory region 1 LR-PCR amplicons originating from BPS eDNA: general observations.** The optimized LR-PCR was used to determine the diversity of the IncP-1 accessory 1 region in matrix samples of BPS Kortrijk taken before, during and after the spraying season. As a first measure of diversity and its seasonal dynamics, the size range of the LR-PCR amplicons was determined by agarose gel electrophoresis. Amplicon sizes ranged between 2.5 and 15 kb and highly similar size profiles were observed between sampling positions and sampling time points (Fig. 1). To guarantee sufficient DNA for sequencing, the eDNAs from all samples taken in July and September were pooled (1:1 volume ratio for all samples) in one composite eDNA sample for high-throughput Illumina sequencing. This is justified since the July and September samples represented samples taken at the height of the spraying season. MetaSPAdes assembled the reads into 333 contigs with lengths ranging between 0.5 and 9.6 kb, yielding 499 kb of unique sequence (Table 1). Using BLASTn, sequences were searched that are expected at the *traC* and *trbP* flanking sides of the amplicons, i.e., *traC* sequences at the *traC* side and *trbP/upf30.5/upf31.0/parA* sequences at the *trbP* side. In total partial *traC* and *trbP/upf30.5/upf31.0/parA* sequences were found on sequence ends of, respectively, 5 (contigs 1, 21, 64, 86, and 190) and 10 contigs (contigs 14, 44, 62, 64, 121, 241, 259, 291, 319, and 327) (Table 1). Contig 64 contained part of *traC* at one end and part of *upf31.0* at the other end indicating that it contained the complete accessory region 1. One contig, i.e., contig 317, appeared to be generated by binding of the trbP_LR primer with the noncanonical *traC* internal trbP_LR binding site since it started with the traC_LR primer site and ended with the *traC* internal trbP_LR binding site. Recovered *traC/trbP/upf30.5/upf31.0/parA* backbone sequences related either to IncP-1$\beta$ (9 contigs), IncP-1$\varepsilon$ (4 contigs) or IncP-1$\delta$ (2 contigs) plasmid subgroups.

**Identification of accessory genes on metagenomic IncP-1 LR-PCR amplicons.** Gene prediction by RAST identified 430 unique ORFs (Table 1), three tRNA sequences and three rRNA sequences. BlastKOALA was used to assign KEGG Orthology (KO) identifiers to the ORFs and to classify the KOs into functional categories (Fig. 2). A KO was assigned to 31.9% (137 ORFs) of the ORFs. However, 46.7% (64 ORFs) of the ORFs that were assigned a KO, were not assigned to a specific functional category. Dominant functional categories identified by BlastKOALA, were amino acid metabolism (14 ORFs), carbohydrate metabolism (13 ORFs), glycan biosynthesis and metabolism (9 ORFs),

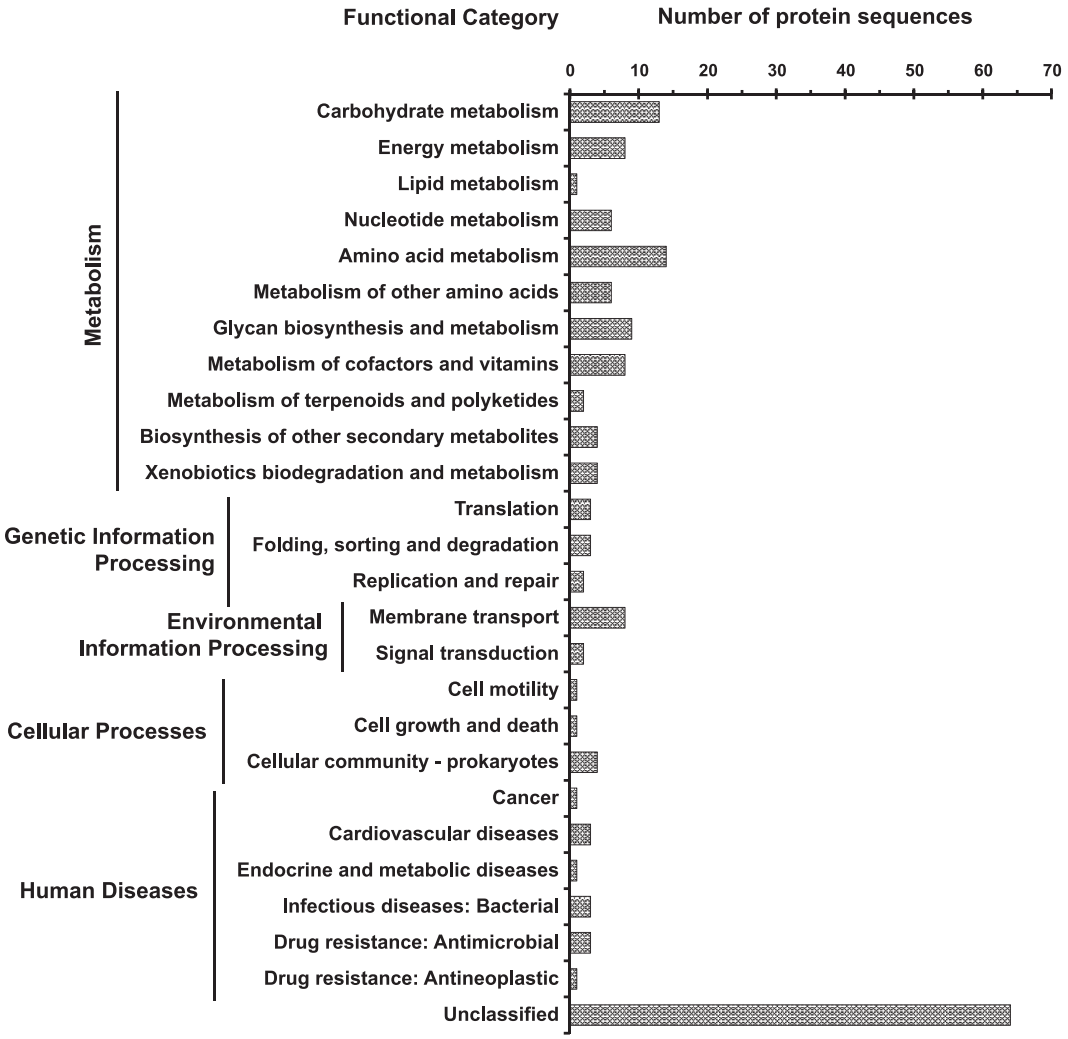


**FIG 2** KEGG orthology (KO) classification of the deduced protein sequences encoded by the ORFs identified on the accessory region 1 amplicons recovered from BPS Kortrijk. BlastKOALA (KEGG Orthology and Links Annotation) was used to assign KOs to the ORFs based on protein similarities using the reconstruct pathway output and to classify ORFs into different functional categories. Numbers represent the number of protein sequences that were assigned to the indicated KO category. In total 31.9% (137 ORFs) of the 430 ORFs were assigned a KO number of which 64 ORFs were not assigned a functional category.

energy metabolism (8 ORFs), metabolism of cofactors and vitamins (8 ORFs), and membrane transport (8 ORFs). The category xenobiotic degradation and metabolism included four ORFs while the antimicrobial drug resistance category contained three ORFs (Fig. 2). Two of the four ORFs assigned to the category xenobiotic degradation and metabolism were assigned exclusively to this category, encoding a putative N-ethylmaleimide reductase (ORF152) and a putative amidohydrolase (ORF308). The other two ORFs encoded a gluconolactonase (ORF126) and a glutathione *S*-transferase (ORF337) and were assigned to different functional categories.

More in-depth analysis of the ORFs was achieved by examining the presence of Hidden Markov Models (HMM) using hmmscan and protein similarities by BLASTp searches against the UniProtKB database. HMM hits were found for 64.9% of the ORFs whereas BLASTp analyses resulted into a hit rate of around 58% independent of the database used (supplementary annotation file). Several recovered functions related to typical accessory traits found on IncP-1 plasmids such as xenobiotic degradation (12 ORFs) and resistance mechanisms (9 ORFs). Xenobiotic degradation functions comprised both upper and lower catabolic pathways (Table 2). Putative upper pathway

**TABLE 2** Overview of ORFs with predicted resistance and organic xenobiotic degradation functions[a]

| ORF | Contig (length bp) | Position | #AA | Coverage | Hmmscan description | BLASTp description | E value | AA-ID (%) | Accession no. |
|---|---|---|---|---|---|---|---|---|---|
| ***Resistance*** | | | | | | | | | |
| 7 | contig001 (9648) | 1817-3463 | 548 | 3289.6 | MerA protein | Mercury (II) reductase | 2E-78 | 96.71 | P06689 |
| 101 | contig017 (3776) | 3752-3492 | 86 | 7.1 | D-ala D-ala ligase C-terminus | D-alanine-D-alanine ligase | 9E-27 | 60.8 | Q83BZ9 |
| 112 | contig022 (316) | 63-923 | 286 | 11.3 | EamA-like transporter family | Protein PecM | 7E-62 | 43.4 | P42194 |
| 113 | contig022 (316) | 2154-955 | 399 | 17.7 | Major Facilitator Superfamily | Tetracycline resistance protein, class A | 0E+00 | 100.0 | P02982 |
| 189 | contig049 (2474) | 179-1654 | 491 | 24.2 | Major Facilitator Superfamily | Antiseptic resistance protein | 6E-76 | 33.9 | P0A0J8 |
| 193 | contig050 (2472) | 1773-2165 | 130 | 7.0 | Glyoxalase/bleomycin resistance protein/ dioxygenase superfamily | Uncharacterized protein Mb0911c | 3E-12 | 32.5 | P64742 |
| 275 | contig086 (1843) | 712-1044 | 110 | 9260.9 | Small Multidrug Resistance protein | Multidrug transporter EmrE | 1E-41 | 62.7 | P23895 |
| 298 | contig102 (1633) | 33-899 | 288 | 25,472.1 | Pterin binding enzyme | Sulfonamide-resistant dihydropteroate synthase type-1 | 0E+00 | 100.0 | P0C002 |
| 346 | contig142 (1245) | 50-445 | 131 | 5.3 | FtsX-like permease family | Macrolide export ATP-binding/permease protein MacB | 9E-27 | 43.5 | Q6MPX9 |
| 425 | contig322 (566) | 162-494 | 110 | 12.2 | Small Multidrug Resistance protein | Quaternary ammonium compound-resistance protein QacF | 3E-66 | 93.6 | Q9X2N9 |
| ***Xenobiotic*** | | | | | | | | | |
| 66 | contig009 (4780) | 3635-4135 | 166 | 5.1 | short chain dehydrogenase | 3-oxoacyl-[acyl-carrier-protein] reductase FabG | 5E-22 | 38.8 | P51831 |
| 72 | contig011 (4547) | 1999-827 | 390 | 10.6 | FAD binding domain | Uncharacterized aromatic compound monooxygenase YhjG | 3E-11 | 24.9 | O07561 |
| 109 | contig020 (3595 bp) | 2517-2981 | 154 | 11.4 | [2Fe-2S] binding domain | Carbon monoxide dehydrogenase small chain | 3E-11 | 24.9 | P19915 |
| 125 | contig026 (3012 bp) | 44-460 | 138 | 5.5 | HpcH/HpaI aldolase/citrate lyase family | 2-keto-3-deoxy-L-rhamnonate aldolase | 1E-16 | 33.9 | B1LLJ9 |
| 162 | contig036 (2732) | 2278-2592 | 104 | 112.6 | Rieske [2Fe-2S] domain | Benzene 1,2-dioxygenase system ferredoxin subunit | 2E-22 | 44.0 | Q07947 |
| 285 | contig092 (1738) | 1568-753 | 271 | 6.8 | short chain dehydrogenase | Hydroxysteroid dehydrogenase-like protein 2 | 3E-103 | 55.0 | Q66KC4 |
| 308 | contig110 (1554) | 984-22 | 320 | 17.5 | Amidohydrolase | 2,3-dihydroxybenzoate decarboxylase | 1E-86 | 45.0 | P80402 |
| 309 | contig110 (1554) | 1537-1013 | 174 | 17.5 | Acyl-CoA dehydrogenase, C-terminal domain | Flavin-dependent monooxygenase, oxygenase subunit HsaA | 1E-21 | 32.3 | Q0S811 |
| 317 | contig117 (1450) | 533-213 | 106 | 9.2 | KDPG and KHG aldolase | 2-dehydro-3-deoxy-phosphogluconate aldolase | 7E-27 | 50.6 | O68283 |
| 344 | contig138 (1268) | 1128-313 | 271 | 5.2 | Oxidoreductase FAD-binding domain | Ferredoxin—NADP reductase | 2E-103 | 58.1 | Q44532 |
| 386 | contig200 (928) | 868-566 | 100 | 2.9 | SnoaL-like domain | | | | |

(Continued on next page)

**TABLE 2** (Continued)

| ORF | Contig (length bp) | Position | #AA | Coverage | Hmmscan description | BLASTp description | E value | AA-ID (%) | Accession no. |
|---|---|---|---|---|---|---|---|---|---|
| 414 | contig263 (669) | 296-439 | 47 | 2.3 | Dienelactone hydrolase family | | | | |
| ***Other*** | | | | | | | | | |
| 109 | contig020 (3595) | 2517-2981 | 154 | 11.4 | [2Fe-2S] binding domain | Carbon monoxide dehydrogenase small chain | 1E-58 | 59.2 | P19915 |
| 280 | contig088 (1825) | 1797-1162 | 211 | 14.3 | Cellulase (glycosyl hydrolase family 5) | Endoglucanase D | 1E-31 | 35.1 | P25472 |

[a]Functions were assigned according to detected Hidden Markov Models (HMM) with hmmscan and BLASTp protein similarities (UniprotKB database) found in the predicted protein sequence of each ORF.

catabolic gene functions encompassed a benzene 1,2-dioxygenase (ORF162), a pyruvate class II aldolase (ORF125), an amidohydrolase (ORF308) and a monooxygenase (ORF72). No prediction of target xenobiotics could be made as AA similarity of these gene functions ranged from 24.9% to 45.0%. In addition, a cellulase function (ORF280; 35.1% AA-ID) was retrieved (Table 2). Predicted resistance gene functions included both metal and antibiotic resistances (Table 2), i.e., resistances to mercury (MerA; ORF5; 99.6% AA-ID), vancomycin (Ddl; ORF101; 60.8% AA-ID), tetracycline (TetA; ORF113; 100% AA-ID), sulfonamide (Sul1; ORF298; 100% AA-ID) and macrolide (ORF346; 64.1% AA-ID). Other recovered resistance functions did not target a specific antibiotic and included bleomycin-resistance like functions (ORF193; 32.5% AA-ID), major facilitator superfamily transporters (ORF189; 33.9% AA-ID) and multidrug resistance transporters (ORF112, ORF275 & ORF425; 43.4–93.6%AA-ID). Most predicted gene products that were annotated as resistance conferring using BLASTp, including resistance to tetracycline (ORF113; *tetA*), sulfonamide (ORF298; *sul1*), mercury (ORF5; *merA*) and a multidrug resistance protein (ORF275 & ORF189), received an unclassified KO number and hence were not retrieved within the antimicrobial drug resistance category of the KEGG analysis. Other often reoccurring gene functions that could be linked with adaptation included transposase functions (9 ORFs) belonging to various insertion sequences families, i.e., IS*1071* (ORF167, ORF402), IS*21* (ORF153, ORF173), IS*6* (ORF83) and IS*4*/IS*5* (ORF367).

Several functions of the recovered genes could not be directly linked to classical adaptive traits resistance and catabolism. A frequently encountered group of genes encoded homologues of ABC transporters (11 ORFs), comprising ORFs that show resemblance to the transportation systems of lipopolysaccharides (54.8–92.2% AA-ID), ribose (35.4–67.6%) and peptides (35.6–67.6% AA-ID). Others encoded homologs of a periplasmic nitrate reductase (contig 27, NapABD, 34.9% AA- 51.4%, AA-ID), a form II aerobic carbon monoxide dehydrogenase (contig 20, CoxLMS, 33.4%–59.2% AA-ID) participating in the Wood-Ljungdahl pathway (24), the SUF system for the biogenesis of iron/sulfur proteins, consisting of two subcomplexes, i.e., SufBCD and SufSE, distributed over several contigs (contigs 23, 36, 58, 127, 50.4% to 70.9% AA-ID), glycosyltransferases (contig 24, 39% AA-ID), proteins involved in lipopolysaccharide (LPS) biosynthesis and transport (contigs 65, 79, 98 and 112, LptA, LptC, KdsC and KdsD, 25.5% to 55.1% AA-ID), and sigma factors $\sigma^{32}$ (contig 96, RpoH, 89.0% AA-ID) and $\sigma^{54}$ (contig 3, RpoN, 61.4% AA-ID) (see also Supplementary Annotation File).

**Genetic organization and synteny of metagenomic LR-PCR amplicons.** To gain insight in the gene synteny of the complete LR amplicons, Mega-BLAST searches were performed against the nt database. Sixteen contigs were identified that showed ≥95% or more nt similarity with a sequence in the NCBI nt-database over a stretch of at least 85% of the complete contig length and containing two or more predicted ORFs (contigs 1, 19, 21, 22, 27, 34, 35, 41, 44, 49, 52, 62, 86, 96, 102 and 121). Contigs 1, 21 and 86 contained the partial IncP-1 plasmid backbone sequence *traC*. The synteny analysis showed that the *merA* homologue in contig 1 was part of a complete predicted mercury resistance gene cluster encompassed in a transposon-like structure identical to Tn*5058* located in the accessory gene regions 1 of IncP-1$\zeta$ plasmid pMCBF1 (100% nt similarity; AY950444) and IncP-1$\alpha$ plasmid pB11 (95% nt similarity; CP002152) (Fig. 3A). In contig 86, the above-mentioned unspecified multidrug resistance function encoded by ORF275, was encompassed in a region that showed 100% nt similarity to the accessory region 1 of IncP-1$\varepsilon$ plasmids pEMT3 and p712.

Other contigs showing significant synteny did not include plasmid backbone genes. The putative tetracycline resistance gene *tetA* (ORF113) present on contig 22 was linked to its regulatory gene *tetR* (ORF114) as well as to the often associated EamA-family transporter gene (ORF112). Contig 102 contained *sul1* (ORF298) encoding sulfonamide resistance and also ORF299, encoding an uncharacterized protein, which often accompanies *sul1*. Interestingly, both the *tet*AR region on contig 22 and the *sul1* region on contig 102 showed synteny with the respective *tetAR* and *sul1* regions found

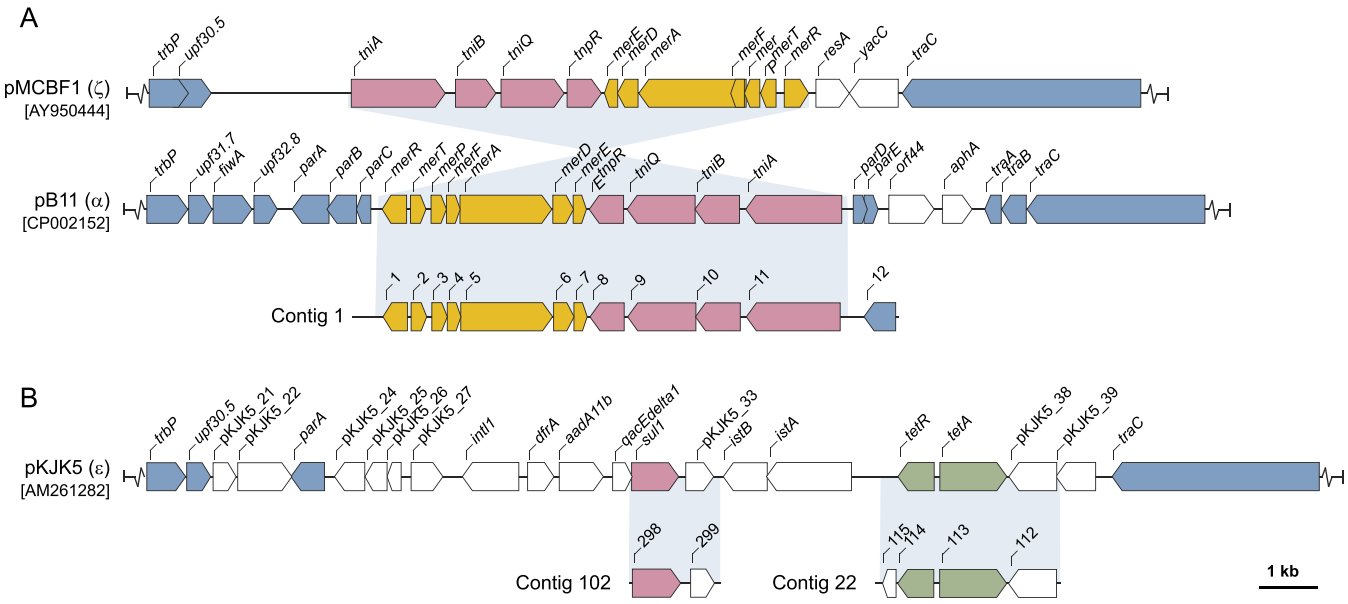


**FIG 3** Schematic representation of the gene synteny of resistance gene clusters retrieved on contigs of IncP-1 accessory region 1 amplicons recovered from BPS Kortrijk. A) On contig 1 a complete Tn*5058* is retrieved, containing a mercury resistance gene cluster (yellow) and transposition machinery (pink), adjacent to a *traC* remnant (ORF12). Synteny with corresponding gene clusters on IncP-1 plasmids are shown by the gray shades (≥95% nt similarity) and IncP-1 backbone genes are represented in blue. B) The resistance mechanisms for sulfonamide (pink) and tetracycline (green), retrieved on, respectively, contig 102 and contig 22, show great resemblance (≥95% nt similarity, gray shades) with resistance gene clusters located on the accessory region 1 of the multiresistance IncP-1$\varepsilon$ plasmid pKJK5. IncP-1 backbone genes are represented in blue.

on other plasmids, including the multidrug resistance IncP-1$\varepsilon$ plasmid pKJK5 where both resistance functions are present together in accessory region 1 (Fig. 3B) (25). Finally, the periplasmatic nitrate reductase *napDAB* gene cluster identified on contig 27 is almost identical (96% nt similarity) to that of the polyaromatic hydrocarbon degrader *Immundisolibacter cernigliae* strain TR3.2 (CP014671). The same *napDAB* organization is present in other bacteria, including *Cupriavidus metallidurans* CH34 (NC_007973).

Synteny was further explored using the BLASTp results of the ORFs in search of conserved gene organization on contigs with less than 85% nt similarity (Supplementary Annotation File). The predicted *coxLMS* operon recovered in contig 20 differed in organization from canonical *cox* operons encoding form I and II CO-dehydrogenases (26) (Fig. S3A) but showed complete synteny with a *cox* operon in an unspecified betaproteobacterium (MERI01000190) isolated from an aquifer system. Similarly, the three consecutive glycosyltransferase genes on contig 24 showed a gene organization that is identical to gene clusters in the borneol degrader *Pseudomonas* sp. TCU-HL1 and in the biphenyl degrader *Pseudomonas pseudoalcaligenes* KF707 (Fig. S3B). In contig 36, the gene cluster predicted to encode SufDS was located upstream of the gene encoding a putative benzene 1,2-dioxygenase that contains a Riekse [2Fe-2S] domain. The region showed substantial synteny with the gene cluster identified in a *Xanthomonadales* bacterium (MIDT01000097) (Fig. S3C).

## DISCUSSION

**LR-PCR targeting IncP-1 cargo regions: opportunities and limitations.** A LR-PCR method that recovers and identifies the accessory region 1 of IncP-1 plasmid subgroups $\beta$, $\varepsilon$ and $\delta$ was successfully developed. In spite of not being all-inclusive for the entire IncP-1 plasmid group, the assay is highly relevant as the three target subgroups constitute the largest fraction of IncP-1 plasmids reported to date and are the most frequently recovered IncP-1 carriers of xenobiotic catabolic gene functions in environmental bacterial isolates (27, 28). We observed some a-specific amplification due to the binding of the trbP_LR primer to sequences near the traC_LR primer binding site but this never interfered with the detection of the actual accessory region 1. Furthermore, the a-specific amplification appears overruled when the canonical

trbP_LR binding sequence is available. The latter is also suggested by the observation that, in agarose gel electrophoresis, none of the BPS samples showed amplicon sizes corresponding to the internal *traC* fragments expected to be generated by binding of the trbP_LR primer to the matching sequence in *traC*. Moreover, due to the high specificity of the traC_LR primer, exclusively regions upstream of *traC* will be amplified hence, including either other IncP-1 backbone genes or parts of the accessory region 1. Therefore, we did not consider this a-specific binding of trbP_LR as an limitation especially since these amplicons will be recognized either by agarose gel electrophoresis or sequencing.

Since our primers were internal to conserved flanking backbone genes, the co-amplification of these genes not only demonstrated that the correct region was amplified but also helped identifying the subgroup of the carrying IncP-1 plasmids. Indeed, the amplified cargo regions originated predominantly from IncP-1$\beta$ subgroup plasmids (9 contigs), followed by IncP-1$\varepsilon$ (4 contigs) and IncP-1$\delta$ (2 contigs) plasmids which is in accordance with Dealtry et al. (8) who demonstrated by *trfA* amplicon sequencing the abundance of several IncP-1 plasmid subgroups in the same BPS matrix, including these three, with the IncP-1$\beta$ plasmids being the most abundant.

Important parameters of the LR-PCR method are the amplicon size and detection limit. Our method enables efficient recovery of amplicons up to 21 kb. Similar maximum sizes were obtained using LR-PCR targeting the cargo of IS*1071* composite transposons (16). Reported IncP-1 accessory regions 1 in data bases are up to 45 kb but 76% of the IncP-1 plasmids ($\beta$, $\varepsilon$ and $\delta$) in the NCBI database carry accessory regions smaller than 21 kb, suggesting that a substantial fraction of the target can be covered. The detection limit of the current LR-PCR method of $10^5$ copies is about 1 order of magnitude higher than LR-PCR targeting the IS*1071* cargo genes (16) and endpoint PCR targeting individual genes (29), but still in the range of reported environmental IncP-1 plasmid abundances (9). As such, the LR-PCR method targeting the IncP-1 accessory region 1 is able to selectively recover long IncP-1 accessory DNA fragments from eDNA with sufficiently high sensitivity. Moreover, amplicon size profiling by agarose gel electrophoresis provides an initial view of the recovered amplicon sizes and thus allows the selection of samples that show unique amplicon sizes of interest for further sequencing.

Overall, the new LR-PCR assay extends the culture-independent mobilome toolbox, with the major benefit that it provides direct insight into the putative functions of accessory genes that are carried by IncP-1 plasmids within a microbial community and hence are prone to HGT as mediated by these highly promiscuous plasmids. The method surpasses other culture-independent methods in unraveling plasmid accessory genes like shotgun or even long-read sequencing of total community DNA because these methods are not able to sequence genes from rare community members without prior selective enrichment. Even sequencing of community-wide plasmid extracts still suffers from poor plasmid extraction efficiencies especially for larger plasmids like those of the IncP-1 group (8, 30, 31). Using long-read sequencing technologies to sequence these amplicons will further improve assembly (32). Limitations include that larger accessory gene clusters are missed and the potential recovery of chimaeras. However, we assume that chimaeras are less likely to occur compared to amplification of highly homologous sequence as in the case of bulk 16S rRNA gene sequence analysis since prematurely terminated amplicons are less likely to reanneal to a different template DNA.

**Limited variation of the IncP-1 accessory region 1 cargo in the BPS environment over the course of an agricultural season.** We have previously examined the dynamics of the bacterial community composition as well as of the IncP-1 abundance/diversity in BPS Kortrijk during the agricultural season of 2011 (but only encompassing samples of March, July and September) (8, 9, 23). A small but clearly reduced diversity in community composition in the samples of July and September compared to those of March, was observed, suggesting effects of the pesticide addition on the BPS community (23). These overall community changes at genus level were, however, minor compared to fluctuations in the composition of IncP-1 conjugative plasmids, which

suggested that HGT is a crucial parameter in community development of an effective pesticide degrading microbial community in the BPS. Indeed, increased numbers of plasmids were observed in July/September along with changes in IncP-1 diversity with the relative abundances of Inc-P-1$\beta$ higher and those of IncP-1$\varepsilon$ lower than the respective numbers and abundances in the March samples (8). The variation in the recovered IncP-1 cargo size profiles and hence of the apparent diversity of IncP-1 cargo was limited during the considered year, suggesting that the cargo size and likely its content did not substantially change. This indicates that the BPS community makes use of an IncP-1 population with an established cargo in accessory region 1. The BPS examined in this study was already operational for 3 years during which a quite stable cargo set in accessory gene region 1 might have established in the community or alternatively, that another IncP-1 region was adaptive.

**Only a minor fraction of the IncP-1 accessory region 1 carries catabolic genes in the examined BPS community.** We hypothesized that the BPS environment with its high pesticide concentrations (16) would select for catabolic genes in accessory region 1 of IncP-1 plasmids. However, while catabolic genes were recovered, their frequency of occurrence was relatively low, i.e., only 3% of the contigs contained catabolic gene homologs. This was far below the 41% previously recovered as IS*1071* cargo from the same BPS environment (16). Hence, the IncP-1 accessory region 1 cargo did not mirror the most obvious selective pressure in the BPS environment, i.e., the continuous input of pesticides. This is in contrast to the IS*1071* cargo and was unexpected since catabolic plasmids derived from isolates from xenobiotic polluted environments often belong to the IncP-1 group and a substantial fraction of those contain catabolic genes in accessory region 1. In our database, all 23 catabolic IncP-1 plasmids belong to the $\beta$, $\varepsilon$ or $\delta$ subgroup and 44% of these carry catabolic genes in this region. No reports exists on catabolic plasmids belonging to the nontarget IncP-1 plasmid groups $\alpha$, $\gamma$, $\zeta$ and $\eta$. Moreover, catabolic genes present on IncP-1 plasmids are often embedded in a composite transposon structure bordered by IS*1071* elements (i.e., 71% of the catabolic cargo genes in our IncP-1 database). The presence of IS*1071* in the IncP-1 accessory region 1 in the BPS community was confirmed through the identification of its partial sequences in two rather small contigs (contigs 39 and 221) containing almost exclusively of IS*1071* sequences, likely due to the repetitive nature of IS*1071* leading to poor assembly. Dunon et al. (16) recovered a large number of IS*1071* composite transposons of less than 21 kb and therefore the observed size limitation of 21 kb of the LR-PCR method is an unlikely explanation for the poor recovery of catabolic genes. Otherwise, the presence of a noncanonical trbB_LR primer sequence in IS*1071* (see Table S2) observed with reference IncP-1$\beta$ plasmids pI2 and pTB30, might have interfered with amplification of complete IS*1071* associated composite transposon containing accessory 1 regions. However, this is unlikely as the sequence matching with the 3′ end trbP_LR primer sequence in IS*1071* was only 8 bp long and did not affect amplification of the complete accessory 1 region containing IS*1071* at a similar position in other plasmids like pNB8c. Moreover, the recovered IS*1071* sequences of contigs 39 and 221 extended largely the IS*1071* fragment that was expected in case amplification occurred from 8 bp TrbP_LR primer matching sequence in IS*1071*. Therefore, IncP-1-$\beta$, -$\varepsilon$ or -$\delta$ plasmids that carry the IS*1071* composite transposons containing catabolic genes in accessory region 1 may only represent a minor fraction in the examined BPS. Interestingly, Dunon et al. (16) found that (i) IS*1071* composite transposons are also present on other plasmids than IncP-1 plasmids and (ii) a third accessory region between *klcA* and *traM*, previously identified in pB4 (33) and currently underrepresented in bacterial isolates, might be another important hot spot of IS*1071* associated catabolic cargo in IncP-1 plasmids in the BPS environment. Apparently, accessory region 1 did not play a major role in the pesticide catabolic adaptation of the BPS community. The lack of catabolic gene functions in accessory region 1 of IncP-1 plasmids does however not necessarily mean that the IncP-1 plasmids are not carrying gene functions involved in pesticide degradation since these genes might be localized in the alternative insertion hot spot, accessory region 2 between *trfA* and *oriV*, or as mentioned above in between *klcA* and *traM*. While the insertion hot spot character of the region between *klcA* and *traM* is not fully clear yet,

several attempts in our research group to develop a suitable primer sets for accessing accessory region 2 between *trfA* and *oriV*, were unsuccessful.

**IncP-1 plasmids are carriers of metal and antibiotic resistance genes in the examined BPS community.** In addition to catabolic genes and in contrast to the observed IS*1071* associated cargo (16), homologues to antibiotic and metal resistance genes were observed as accessory region 1 cargo of IncP-1 plasmids in the examined BPS environment. Several of these demonstrated much higher sequencing coverages (up to 150–fold) than the catabolic IncP-1 cargo and therefore are likely present at higher relative abundances than the catabolic functions (Table 2). Among the recovered predicted resistance gene clusters was the *mer* cluster conferring mercury resistance. Mercury resistance has been previously associated with IncP-1 plasmids with no preference for one of the two major IncP-1 insertion hot spots (34–36). *mer* clusters have also been linked to plasmids in the BPS Kortrijk environment. Those plasmids were obtained from isolates from BPS Kortrijk (37) or by exogenous isolation (22). As shown before, the occurrence of mercury resistance within the BPS community cannot be directly related to a selection pressure (34, 35). The mercury resistance may very well hitchhike with catabolic genes present in the other cargo insertion sites on the plasmid. IncP-1-$\beta$ and -$\varepsilon$ plasmids that carry both mercury resistance and catabolic genes have been reported with either the *mer* cluster inserted in accessory region 2 and the catabolic genes in accessory region 1, or both determinants nested in accessory region 2 (28, 36).

In addition to mercury resistance gene homologues, several predicted antibiotic resistance genes (ARG) could be recovered. The most obvious ones included the primarily clinically associated ARGs encoding sulfonamide (*sul1*) and tetracycline (*tetAR*) resistance. The *sul1* and *tetAR* genes were previously reported as cargo of either IncP-1$\beta$ or IncP-1$\varepsilon$ plasmids (38–41). Several IncP-1$\varepsilon$ plasmids carry both *tetAR* and *sul1* nested together in accessory region 1 (39) while in the IncP-1$\beta$ plasmid pB10, *sul1* is located within the *intI1* integron in accessory region 1 and *tetAR* in accessory region 2 (40). Several other ARGs were identified, although with less similarity, including a macrolide antibiotic resistance. To date, IncP-1$\beta$ plasmid pRSB111 from a wastewater treatment plant isolate is the only plasmid that carries a macrolide resistance module, also in accessory region 1 (42). Our findings are consistent with the previous detection of these antibiotic resistance genes in BPS Kortrijk either by conventional PCR on eDNA or in plasmid extracts from isolates (8, 37), including *sul1* and *tetAR* but also others like *aadA* and *sul2* (8, 43). The occurrence of ARGs in BPS Kortrijk defined the BPS environment as a possible reservoir of ARGs (8, 43). We here show that at least some of these ARGs are carried by IncP-1 plasmids and are therefore prone to HGT. BPSs should thus be added to the list of potential reservoirs of promiscuous ARGs outside clinical settings, joining soil, manure, wastewater treatment plants and biogas plant digestates (28, 39, 41, 44). Since horse manure was added at the start of the spraying season as a nutrient source in BPS Kortrijk, it may have introduced both a source, i.e., bacteria carrying ARGs, and the selection pressure, i.e., antibiotic compounds for plasmid borne ARGs as previously suggested for manure amended soil (8). However, according to the information provided by the operators of the examined BPS, antibiotics were not present in the BPS influent water and hence might not compose a selective force. Therefore, maintenance of ARGs in the BPS environment might be explained otherwise than by direct selection by antibiotics. ARGs that determine export systems (like *tetA*) might be involved in tolerance to toxic pesticides (45–47) or they might be selected as a protective response to naturally produced antibiotics in the BPS environment, where they might function as toxins or effectors (48). Otherwise, consistent with the *mer* homologues, coselection of antibiotic resistances in accessory region 1 may have occurred due to the physical linkage with catabolic genes located elsewhere on the plasmids. Manure is not a frequently used ingredient of the BPS matrix and it remains to be seen if communities of other BPSs carry (IncP-1 plasmid encoded) ARGs. Albeit, their occurrence and mobile character in BPS Kortrijk emphasizes the need for careful composition of the BPS matrix at startup, the safe operation of the BPS as well as the

safe disposal of the spent matrix for reasons beyond the potentially high residual pesticide content (18).

**IncP-1 cargo includes a variety of gene functions not directly related to the known selective pressures of BPSs.** A diverse range of IncP-1 accessory region 1 associated gene functions were predicted that were not the typical adaptive trait determinants like resistance and catabolism. The advantageous role they play in host fitness is currently unclear and remains speculative, especially since % AA identity with known functions did often not exceed 60%. Moreover, a substantial number of ORFs could not be assigned a function. Sentchilo et al. (30) also recovered plasmid genes encoding nontypical accessory functions such as membrane transport and stress-relief gene functions as well as a large number of unknown genes, when sequencing plasmids directedly extracted from wastewater sludge while Shintani et al. (2020) observed a wide variety of accessory gene functions, including unknown ones, in IncP-1 plasmids exogenously isolated from the rhizosphere of lettuce and tomato plants grown in unpolluted soil (49). Similarly, the IS*1071* accessory genes identified by Dunon et al. (16) in the BPS included 30–40% of unknown gene functions and various gene functions that were not obviously adaptive. A possible explanation is that these traits are involved in regulating and optimizing catabolic gene functions or assist the cells in coping with direct or indirect stress exerted by the pollution. For instance, the recovered gene functions predicted to be involved in the biogenesis of iron/sulfur proteins (*suf*-gene functions) could aid in the synthesis of xenobiotic degradative mono- and dioxygenases containing the iron-sulfur Rieske [2Fe-2S] domain in their catalytic site (50). Similarly, predicted ABC transporters might be involved in uptake of xenobiotics for degradation or otherwise in their export to reduce toxicity (51). Alternatively, other unknown environmental factors beyond the pesticides may select for these traits. For instance, the putative glycosyltransferases might take part in the functionalisation of the lipopolysaccharides in the outer membrane of Gram-negative bacteria, which in turn would contribute to resistance to antimicrobial agents produced by other organisms in the BPS environment (52). Similarly, the predicted periplasmic nitrate reductase (*napDAB*-operon) was initially identified for its role in prokaryotic anaerobic nitrate respiration, but its function is diverse, including nitrate scavenging and pathogenicity, which might create competitive advantages in the BPS environment (53). Alternatively, as suggested for the *mer* and ARG homologues, these functions may just hitchhike with catabolic genes elsewhere on the plasmid. Finally, it is also possible that during their transfer among bacteria, IncP-1 plasmids acquire non-host-beneficial genes at sufficiently high frequencies for us to detect them during this 'snapshot' in time and hence that these genes are in fact remnants of random recruitment. Whatever the explanation, the wide variety of recovered cargo gene functions, including atypical adaptive trait determinants and unknown functions, emphasizes the role of IncP-1 plasmids in bacterial community diversification in the BPS environment and shows the potency of the LR-PCR method in acquiring new information on IncP-1 plasmid ecology in complex microbial communities.

## MATERIALS AND METHODS

**Plasmids and plasmid extracts used in this study.** Plasmids used in this study and their host species are presented in Supplementary information Table 3. Host strain cultivation and plasmid extraction from the bacterial cultures were performed as previously described (16).

**BPS samples and eDNA.** The high molecular weight (HMW) eDNA extracts used for long-range PCR were those that were previously reported in Dunon et al. (16) The extracts originated from biomix samples taken from a BPS operating since 2008 at an agricultural farm in Kortrijk, Belgium. Detailed information about the sampling procedure of the biomix is found in Dealtry et al. (8) Briefly, samples were collected before the annual startup of the BPS system in March 2011, during the spraying season in July and September 2011 and 2 weeks after closing operation in December 2011. Sampling was performed by dividing the BPS (length 20 m, width 1.2 m, depth 1 m) into four sampling sections of 5 m distance and in each section, 12 biomix cores (Ø4 cm; until 10 cm depth) were randomly collected. The cores were mixed per section and the mixed material was sieved (2 mm pore size), resulting as such in 4 composite samples per time point, i.e., 16 composite samples in total. The HMW eDNA originated from 300 mg wet weight material (16). A complete list of all active substances used on the farm in Kortrijk during the agricultural season of 2011 is given in Table S3.

**Long-range PCR.** A data set of 63 reference full IncP-1 plasmid sequences, including eight IncP-1$\alpha$, 39 IncP-1$\beta$, three IncP-1$\gamma$, three IncP-1$\delta$ and 10 IncP-1$\varepsilon$ plasmids, was collected from the NCBI database

**TABLE 3** IncP-1 plasmids used in this study

| Plasmid | Host[a] | Size (kb) | *trbP-traC* amplicon size (bp) | Relevant characteristics | Plasmid sequence accession no. |
|---|---|---|---|---|---|
| IncP-1$\alpha$ | | | | | |
| RP4 | *Escherichia coli* | 60 | - | $Ap^R$; $Km^R$; $Tc^R$ | L27758 |
| pSP21 | *Escherichia coli* | 73 | - | $Km^R$; $Tc^R$ | CP002153 |
| IncP-1$\beta$ | | | | | |
| R751 | *Escherichia coli* | 53 | 9807 | $Tc^R$; $Km^R$ | U67194 |
| pNB8c | *Escherichia coli* | 60 | 20888 | 3,4-$DCA^C$ | JF274990 |
| pUO1 | *Delftia acidovorans* | 67 | 13149 | $Haloacetate^C$ | AB063332 |
| pAKD18 | *Pseudomonas putida* | 67 | 3433 | $Hg^R$ | JN106169 |
| pTB30 | *Escherichia coli* | 79 | 39756 | 3-$CA^C$ | JF274987 |
| pI2 | *Escherichia coli* | 84 | 30601 | 3-$CA^C$ | JF274989 |
| pADP-1::Tn*5* | *Escherichia coli* | 109 | - | $Km^R$; $Atrazine^C$ | U66917 |
| IncP-1$\delta$ | | | | | |
| pAKD4 | *Pseudomonas putida* | | 12351 | $Hg^R$ | GQ983559 |
| pEST4011 | *Alcaligenes xylosoxidans* | 77 | - | 2,4-$D^C$, 3-$CB^C$ | AY540995 |
| IncP-1$\varepsilon$ | | | | | |
| pAKD16 | *Escherichia coli* | | 12892 | $Hg^R$ | JN106167 |
| pKJK5 | *Escherichia coli* | 54 | 17257 | $Tc^R$ | AM261282 |
| pEMT3 | *Escherichia coli* | 63 | 3526 | $Km^R$, 2,4-$D^C$, 3-$CB^C$ | JX469827 |
| IncP-1$\gamma$ | | | | | |
| pMBUI1a | *Escherichia coli* | 44 | - | | JQ432563 |
| pQKH54 | *Escherichia coli* | 70 | - | $Hg^R$ | AJ744860 |

[a]Hosts are the bacteria in which the plasmids were propagated for extraction and not necessarily the original hosts. Ap, ampicillin; Km, kanamycin; Tc, tetracycline; $T_m$, trimethoprim; 3,4-DCA, 3,4-dichloroaniline; 3-CA, 3-chloroaniline; 2,4-D, 2,4-dichlorophenoxyacetic acid; 3-CB, 3-chlorobenzoate; $^R$: resistance gene cluster; $^C$: catabolic gene cluster.

(see Supplementary Information). Forward primer and reverse primer for LR-PCR targeting accessory gene region 1 were designed within *traC* and *trbP* and were designated traC_LR (5′-GRCGATGCARTA CATCCAGGAGGACGGCAC-3′) and trbP_LR (5′-CBTTCTAYGCCTATTACCCGCTGCACCTGG-3′), respectively (Fig. S1). The *traC* and *trbP* sequences were aligned with CLUSTAL X (54) and degenerate primers were manually designed. The LR-PCR was performed in 25 $\mu$l containing 1 U Long PCR enzyme mix (Thermo Scientific, Germany) and 1× Long PCR buffer, 200 $\mu$M each dNTP, 0.1 mg $ml^{-1}$ BSA, 4% (vol/vol) DMSO and 0.4 $\mu$M primers traC_LR and trbP_LR, and template DNA. Amount of used template DNA was around 100 ng per reaction in case of plasmid DNA and 20–30 ng in case of eDNA. The PCR included an initial denaturation of 2 min at 94°C, 10 cycles of 20 s at 94°C and elongation of 16 min at 68°C, followed by 20–25 cycles in which the elongation time extended with 5 s every cycle and ending with an elongation step of 10 min at 68°C. PCR products were separated by agarose gel electrophoresis (0.8%) in Tris-acetate/EDTA buffer and visualized using 1× GelRed™ (Biotium).

**Southern blot hybridization.** A 10 $\mu$l aliquot of LR-PCR products was digested with 1U of restriction enzyme AhdI (Thermo Scientific, Germany) according to the manufacturer's recommendations. The restriction fragments were blotted and hybridized using a *dcaQ* probe as previously described (16).

***De novo* sequencing of LR-PCR amplicons and annotation.** LR-PCR amplicons were ethanol precipitated in 10 mM Tris-HCl/EDTA [pH 8.0]. High-throughput sequencing was performed by BGI Hong Kong, based on an insert library of 500 bp and paired-end sequencing on the HiSeq 2000 (Illumina) platform. The online data platform KBase (55) was used for performing the read assembly with metaSPAdes v1.1.1 (56). Gene prediction and annotation was performed by combining RAST (57), HMMER (58), BLASTp (59) and MegaBLAST (59), using the best hit approach for similarity inference. KEGG Orthologies (KOs) were assigned by BlastKOALA (KEGG Orthology and Links Annotation), using the "species_prokaryotes" database, and classification in functional categories was performed using the reconstruct pathway output (60). Synteny was identified using MegaBLAST (59) on the complete contigs against the nt database, for highly similar gene organizations, and BLASTp on the ORFs sequences for gene organization with nt similarities $\geq$ 85%.

**Data availability.** Raw sequences are deposited in the Sequence Read Archive under the BioProject PRJNA278476. Assembled and annotated contigs are deposited in the GenBank database with accession numbers MK843322 to MK843654.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**SUPPLEMENTAL FILE 1**, PDF file, 1.6 MB.

**SUPPLEMENTAL FILE 2**, XLSX file, 0.3 MB.

## ACKNOWLEDGMENTS

This research was supported by IWT-Vlaanderen Strategic Basic Research project 91370 and the EU project METAEXPLORE (EU grant n°222625). We thank I. Veltens (KH Leuven), for technical assistance, E. Krögerrecklenfort (JKI, Braunschweig) and D. Sen (University of Idaho) for providing IncP-1 plasmids and I. Vercruysse for providing BPS samples and information about the operation of BPS Kortrijk.

## REFERENCES

1. Frost LS, Leplae R, Summers AO, Toussaint A. 2005. Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol 3:722–732. https://doi.org/10.1038/nrmicro1235.
2. Iwasaki W, Takagi T. 2009. Rapid pathway evolution facilitated by horizontal gene transfers across prokaryotic lineages. PLoS Genet 5:e1000402. https://doi.org/10.1371/journal.pgen.1000402.
3. Thomas CM. 2000. Paradigms of plasmid organization. Mol Microbiol 37:485–491. https://doi.org/10.1046/j.1365-2958.2000.02006.x.
4. Shintani M, Takahashi Y, Yamane H, Nojiri H. 2010. The behavior and significance of degradative plasmids belonging to Inc groups in *Pseudomonas* within natural environments and microcosms. Microbes Environ 25:253–265. https://doi.org/10.1264/jsme2.me10155.
5. Brown CJ, Sen D, Yano H, Bauer ML, Rogers LM, Van der Auwera GA, Top EM. 2013. Diverse broad-host-range plasmids from freshwater carry few accessory genes. Appl Environ Microbiol 79:7684–7695. https://doi.org/10.1128/AEM.02252-13.
6. Top EM, Springael D, Boon N. 2002. Catabolic mobile genetic elements and their potential use in bioaugmentation of polluted soils and waters. FEMS Microbiol Ecol 42:199–208. https://doi.org/10.1111/j.1574-6941.2002.tb01009.x.
7. Dennis JJ. 2005. The evolution of IncP catabolic plasmids. Curr Opin Biotechnol 16:291–298. https://doi.org/10.1016/j.copbio.2005.04.002.
8. Dealtry S, Holmsgaard PN, Dunon V, Jechalke S, Ding G-C, Krögerrecklenfort E, Heuer H, Hansen LH, Springael D, Zühlke S, Sørensen SJ, Smalla K. 2014. Shifts in abundance and diversity of mobile genetic elements after the introduction of diverse pesticides into an on-farm biopurification system over the course of a year. Appl Environ Microbiol 80:4012–4020. https://doi.org/10.1128/AEM.04016-13.
9. Jechalke S, Dealtry S, Smalla K, Heuer H. 2013. Quantification of IncP-1 plasmid prevalence in environmental: samples. Appl Environ Microbiol 79:1410–1413. https://doi.org/10.1128/AEM.03728-12.
10. Sen D, Brown CJ, Top EM, Sullivan J. 2013. Inferring the evolutionary history of IncP-1 plasmids despite incongruence among backbone gene trees. Mol Biol Evol 30:154–166. https://doi.org/10.1093/molbev/mss210.
11. Zhang M, Warmink J, Pereira e Silva MC, Brons J, Smalla K, van Elsas JD. 2015. IncP-1$\beta$ Plasmids are important carriers of fitness traits for *Variovorax* species in the mycosphere: two novel plasmids, pHB44 and pBS64, with differential effects unveiled. Microb Ecol 70:141–153. https://doi.org/10.1007/s00248-014-0550-y.
12. Zhang M, Brons JK, van Elsas JD. 2016. The complete sequences and ecological roles of two IncP-1ß plasmids, pHB44 and pBS64, isolated from the mycosphere of *Laccaria proxima*. Front Microbiol 7:909. https://doi.org/10.3389/fmicb.2016.00909.
13. Rios Miguel AB, Jetten MSM, Welte CU. 2020. The role of mobile genetic elements in organic micropollutant degradation during biological wastewater treatment. Water Res X 9:100065. https://doi.org/10.1016/j.wroa.2020.100065.
14. Brown Kav A, Benhar I, Mizrahi I. 2013. A method for purifying high quality and high yield plasmid DNA for metagenomic and deep sequencing approaches. J Microbiol Methods 95:272–279. https://doi.org/10.1016/j.mimet.2013.09.008.
15. Jia H, Guo Y, Zhao W, Wang K. 2014. Long-range PCR in next-generation sequencing: comparison of six enzymes and evaluation on the MiSeq sequencer. Sci Rep 4:5737–5738. https://doi.org/10.1038/srep05737.
16. Dunon V, Bers K, Lavigne R, Top EM, Springael D. 2018. Targeted metagenomics demonstrates the ecological role of IS*1071* in bacterial community adaptation to pesticide degradation. Environ Microbiol 20:4091–4111. https://doi.org/10.1111/1462-2920.14404.
17. Sota M, Tsuda M, Yano H, Suzuki H, Forney LJ, Top EM. 2007. Region-specific insertion of transposons in combination with selection for high plasmid transferability and stability accounts for the structural similarity of IncP-1 plasmids. J Bacteriol 189:3091–3098. https://doi.org/10.1128/JB.01906-06.
18. Karanasios E, Tsiropoulos NG, Karpouzas DG. 2012. On-farm biopurification systems for the depuration of pesticide wastewaters: recent biotechnological advances and future perspectives. Biodegradation 23:787–802. https://doi.org/10.1007/s10532-012-9571-8.
19. Vandermaesen J, Horemans B, Bers K, Vandermeeren P, Herrmann S, Sekhar A, Seuntjens P, Springael D. 2016. Application of biodegradation in mitigating and remediating pesticide contamination of freshwater resources: state of the art and challenges for optimization. Appl Microbiol Biotechnol 100:7361–7376. https://doi.org/10.1007/s00253-016-7709-z.
20. Dunon V, Sniegowski K, Bers K, Lavigne R, Smalla K, Springael D. 2013. High prevalence of IncP-1 plasmids and IS*1071* insertion sequences in on-farm biopurification systems and other pesticide-polluted environments. FEMS Microbiol Ecol 86:415–431. https://doi.org/10.1111/1574-6941.12173.
21. Dealtry S, Ding GC, Weichelt V, Dunon V, Schlüter A, Martini MC, Del Papa MF, Lagares A, Amos GCA, Wellington EMH, Gaze WH, Sipkema D, Sjöling S, Springael D, Heuer H, Van Elsas JD, Thomas C, Smalla K. 2014. Cultivation-independent screening revealed hot spots of IncP-1, IncP-7 and IncP-9 plasmid occurrence in different environmental habitats. PLoS One 9:e89922. https://doi.org/10.1371/journal.pone.0089922.
22. Dealtry S, Nour EH, Holmsgaard PN, Ding G-C, Weichelt V, Dunon V, Heuer H, Hansen LH, Sørensen SJ, Springael D, Smalla K. 2016. Exploring the complex response to linuron of bacterial communities from biopurification systems by means of cultivation-independent methods. FEMS Microbiol Ecol 92:fiv157. https://doi.org/10.1093/femsec/fiv157.
23. Holmsgaard PN, Dealtry S, Dunon V, Heuer H, Hansen LH, Springael D, Smalla K, Riber L, Sørensen SJ. 2017. Response of the bacterial community in an on-farm biopurification system, to which diverse pesticides are introduced over an agricultural season. Environ Pollut 229:854–862. https://doi.org/10.1016/j.envpol.2017.07.026.
24. Janssen PJ, van Houdt R, Moors H, Monsieurs P, Morin N, Michaux A, Benotmane MA, Leys N, Vallaeys T, Lapidus A, Monchy S, Médigue C, Taghavi S, McCorkle S, Dunn J, van der Lelie D, Mergeay M. 2010. The complete genome sequence of *Cupriavidus metallidurans* strain CH34, a master survivalist in harsh and anthropogenic environments. PLoS One 5:e10433. https://doi.org/10.1371/journal.pone.0010433.
25. Bahl MI, Hansen LH, Goesmann A, Sørensen SJ. 2007. The multiple antibiotic resistance IncP-1 plasmid pKJK5 isolated from a soil environment is phylogenetically divergent from members of the previously established $\alpha$, $\beta$ and $\delta$ sub-groups. Plasmid 58:31–43. https://doi.org/10.1016/j.plasmid.2006.11.007.
26. King GM, Weber CF. 2007. Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. Nat Rev Microbiol 5:107–118. https://doi.org/10.1038/nrmicro1595.
27. Poh RPC, Smith ARW, Bruce IJ. 2002. Complete characterisation of Tn*5530* from *Burkholderia cepacia* strain 2a (pIJB1) and studies of 2,4-dichlorophenoxyacetate uptake by the organism. Plasmid 48:1–12. https://doi.org/10.1016/S0147-619X(02)00018-5.
28. Sen D, Van der Auwera GA, Rogers LM, Thomas CM, Brown CJ, Top EM. 2011. Broad-host-range plasmids from agricultural soils have IncP-1 backbones with diverse accessory genes. Appl Environ Microbiol 77:7975–7983. https://doi.org/10.1128/AEM.05439-11.
29. Leys NMEJ, Ryngaert A, Bastiaens L, Verstraete W, Top EM, Springael D. 2004. Occurrence and phylogenetic diversity of *Sphingomonas* strains in soils contaminated with polycyclic aromatic hydrocarbons. Appl Environ Microbiol 70:1944–1955. https://doi.org/10.1128/AEM.70.4.1944-1955.2004.
30. Sentchilo V, Mayer AP, Guy L, Miyazaki R, Tringe SG, Barry K, Malfatti S, Goessmann A, Robinson-Rechavi M, Van Der Meer JR. 2013. Community-wide plasmid gene mobilization and selection. ISME J 7:1173–1186. https://doi.org/10.1038/ismej.2013.13.
31. Li A-D, Li L-G, Zhang T. 2015. Exploring antibiotic resistance genes and metal resistance genes in plasmid metagenomes from wastewater treatment plants. Front Microbiol 6:1025. https://doi.org/10.3389/fmicb.2015.01025.

32. Che Y, Xia Y, Liu L, Li AD, Yang Y, Zhang T. 2019. Mobile antibiotic resistome in wastewater treatment plants revealed by Nanopore metagenomic sequencing. Microbiome 7:1–13. https://doi.org/10.1186/s40168-019-0663-0.
33. Tauch A, Schlüter A, Bischoff N, Goesmann A, Meyer F, Pühler A. 2003. The 79,370-bp conjugative plasmid pB4 consists of an IncP-1$\beta$ backbone loaded with a chromate resistance transposon, the *strA-strB* streptomycin resistance gene pair, the oxacillinase gene $bla_{(NPS-1)}$, and a tripartite antibiotic efflux system of the resistance-nodulation-division family. Mol Genet Genomics 268:570–584. https://doi.org/10.1007/s00438-002-0785-z.
34. Drønen AK, Torsvik V, Goksøyr J, Top EM. 1998. Effect of mercury addition on plasmid incidence and gene mobilizing capacity in bulk soil. FEMS Microbiol Ecol 27:381–394. https://doi.org/10.1016/S0168-6496(98)00085-3.
35. Drønen AK, Torsvik V, Top EM. 1999. Comparison of the plasmid types obtained by two distantly related recipients in biparental exogenous plasmid isolations from soil. FEMS Microbiol Lett 176:105–110. https://doi.org/10.1016/S0378-1097(99)00187-1.
36. Boon N, Goris J, De Vos P, Verstraete W, Top EM. 2001. Genetic diversity among 3-chloroaniline- and aniline-degrading strains of the *Comamonadaceae*. Appl Environ Microbiol 67:1107–1115. https://doi.org/10.1128/AEM.67.3.1107-1115.2001.
37. Martini MC, Wibberg D, Lozano M, Torres Tejerizo G, Albicoro FJ, Jaenicke S, Van Elsas JD, Petroni A, Garcillan-Barcia MP, De La Cruz F, Schlüter A, Pühler A, Pistorio M, Lagares A, Del Papa MF. 2016. Genomics of high molecular weight plasmids isolated from an on-farm biopurification system. Sci Rep 6:28284. https://doi.org/10.1038/srep28284.
38. Popowska M, Krawczyk-Balska A. 2013. Broad-host-range IncP-1 plasmids and their resistance potential. Front Microbiol 4:44. https://doi.org/10.3389/fmicb.2013.00044.
39. Heuer H, Binh CTT, Jechalke S, Kopmann C, Zimmerling U, Krögerrecklenfort E, Ledger T, González B, Top E, Smalla K. 2012. IncP-1$\varepsilon$ Plasmids are important vectors of antibiotic resistance genes in agricultural systems: diversification driven by class 1 integron gene cassettes. Front Microbiol 3:2. https://doi.org/10.3389/fmicb.2012.00002.
40. Schlüter A, Heuer H, Szczepanowski R, Forney LJ, Thomas CM, Pühler A, Top EM. 2003. The 64 508 bp IncP-1 $\beta$ antibiotic multiresistance plasmid pB10 isolated from a waste-water treatment plant provides evidence for recombination between members of different branches of the IncP-1 $\beta$ group. Microbiology (Reading) 149:3139–3153. https://doi.org/10.1099/mic.0.26570-0.
41. Wolters B, Kyselková M, Krögerrecklenfort E, Kreuzig R, Smalla K. 2014. Transferable antibiotic resistance plasmids 1 from biogas plant digestates often belong to the IncP-1$\varepsilon$ subgroup. Front Microbiol 5:765. https://doi.org/10.3389/fmicb.2014.00765.
42. Szczepanowski R, Krahn I, Bohn N, Pühler A, Schlüter A. 2007. Novel macrolide resistance module carried by the IncP-1$\beta$ resistance plasmid pRSB111, isolated from a wastewater treatment plant. Antimicrob Agents Chemother 51:673–678. https://doi.org/10.1128/AAC.00802-06.
43. Martini MC, Quiroga MP, Pistorio M, Lagares A, Centrón D, Del Papa MF. 2018. Novel environmental class 1 integrons and cassette arrays recovered from an on-farm bio-purification plant. FEMS Microbiol Ecol 94:1–9.
44. Schlüter A, Szczepanowski R, Pühler A, Top EM. 2007. Genomics of IncP-1 antibiotic resistance plasmids isolated from wastewater treatment plants provides evidence for a widely accessible drug resistance gene pool. FEMS Microbiol Rev 31:449–477. https://doi.org/10.1111/j.1574-6976.2007.00074.x.
45. Blanco P, Hernando-Amado S, Reales-Calderon J, Corona F, Lira F, Alcalde-Rico M, Bernardini A, Sanchez M, Martinez J. 2016. Bacterial multidrug efflux pumps: much more than antibiotic resistance determinants. Microorganisms 4:14. https://doi.org/10.3390/microorganisms4010014.
46. Muller JF, Stevens AM, Craig J, Love NG. 2007. Transcriptome analysis reveals that multidrug efflux genes are upregulated to protect *Pseudomonas aeruginosa* from pentachlorophenol stress. Appl Environ Microbiol 73:4550–4558. https://doi.org/10.1128/AEM.00169-07.
47. Ghosh S, Cremers CM, Jakob U, Love NG. 2011. Chlorinated phenols control the expression of the multidrug resistance efflux pump MexAB-OprM in *Pseudomonas aeruginosa* by interacting with NalC. Mol Microbiol 79:1547–1556. https://doi.org/10.1111/j.1365-2958.2011.07544.x.
48. Bengtsson-Palme J, Kristiansson E, Larsson DGJ. 2018. Environmental factors influencing the development and spread of antibiotic resistance. FEMS Microbiol Rev 42:fux053. https://doi.org/10.1093/femsre/fux053.
49. Shintani M, Nour E, Elsayed T, Blau K, Wall I, Jechalke S, Spröer C, Bunk B, Overmann J, Smalla K. 2020. Plant species-dependent increased abundance and diversity of IncP-1 plasmids in the rhizosphere: new insights into their role and ecology. Front Microbiol 11:590776. https://doi.org/10.3389/fmicb.2020.590776.
50. Duarte M, Jauregui R, Vilchez-Vargas R, Junca H, Pieper DH. 2014. AromaDeg, a novel database for phylogenomics of aerobic bacterial degradation of aromatics. Database (Oxford) 2014:bau118. https://doi.org/10.1093/database/bau118.
51. Wilkens S. 2015. Structure and mechanism of ABC transporters. F1000Prime Rep 7:14–19. https://doi.org/10.12703/P7-14.
52. Cote JM, Taylor EA. 2017. The glycosyltransferases of LPS core: a review of four heptosyltransferase enzymes in context. Int J Mol Sci 18. https://doi.org/10.3390/ijms18112256.
53. Sparacino-Watkins C, Stolz JF, Basu P. 2014. Nitrate and periplasmic nitrate reductases. Chem Soc Rev 43:676–706. https://doi.org/10.1039/c3cs60249d.
54. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948. https://doi.org/10.1093/bioinformatics/btm404.
55. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia JM, Chia JM, Colasanti R, Conrad N, Davis JJ, Davison BH, Dejongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, Joachimiak MP, Keegan KP, Kondo S, Kumar V, et al. 2018. KBase: the United States department of energy systems biology knowledgebase. Nat Biotechnol 36:566–569. https://doi.org/10.1038/nbt.4163.
56. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. MetaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824–834. https://doi.org/10.1101/gr.213959.116.
57. Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.
58. Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res 39:W29–37. https://doi.org/10.1093/nar/gkr367.
59. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389.
60. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428:726–731. https://doi.org/10.1016/j.jmb.2015.11.006.